June 1947

BIOMETRICS

Vol. 3 NO. 2

THE BIOMETRICS SECTION, AMERICAN STATISTICAL ASSOCIATION

THE ANALYSIS OF COVARIANCE METHOD FOR THE RELATION BETWEEN A PART AND THE WHOLE

R. A. FISHER

University of Cambridge

At the suggestion of Dr. C. I. Bliss and by the courtesy of Professor H. G. O. Holck, whose data I shall use, the following note may serve to illustrate the extreme simplicity with which the technique derived from the analysis of covariance may be applied to problems concerned with the relation of a part to the whole, such as are constantly arising in many fields.

The data consist of the body weights in kilograms and the heart weights in grams of 144 cats used in a group of digitalis assays.[1] Of these 47 were females and 97 males. These data are presented in the table at the end of this note. To simplify the calculations only one decimal place was used for each value. Thus we have:

TABLE 1

TOTAL WEIGHTS

	Females	*Males*
Number	47	97
Total body weight	110.9 Kg.	281.3 Kg.
Total heart weight	432.5 g.	1098.2 g.
Heart as fraction of entire body	.3900%	.3904%

The observed variation in these two measurements can, of course, be expressed by means of the sums of squares and products, as in the following tables. The rather intimidating phrase "spurious correlation" used in the earlier literature sometimes prevents workers from taking the simplest course. Obviously it would be easy to derive from

1 Holck, Harald G. O., Kazuo K. Kimura, and Barbara Bartels, "Effect of the Anesthetic and the Rate of Injection of Digitalis upon Its Lethal Dose in Cats," *Journal of the American Pharmaceutical Assn. 35:* 366–370 (1946).

the crude figures the corresponding square and product for the difference between our variates, representing the weight of the body less the weight of the heart.

TABLE 2

SUMS OF SQUARES AND PRODUCTS FOR BODY AND HEART WEIGHTS

Females	*d.f.*	*(Body)*^{2}	*(Body · Heart)*	*(Heart)*^{2}
Total	47	265.13	1029.62	4064.71
Correction for Mean	1	261.677	1020.516	3979.920
Difference	46	3.453	9.104	84.790
Males				
Total	97	836.75	3275.55	13056.17
Correction for Mean	1	815.77	3185.07	12435.700
Difference	96	20.98	90.48	620.470

From the corrected sums of squares and products we may find the regressions of heart weight on body weight, namely .2637% for females and .4313% for males. It will be noticed that these values are in the first case less and in the second case more than the average contribution of the heart to the total weight. The significance of such differences is often of importance, and must often appear to present a rather complex problem.

We may, however, recognize in the above table the requisite data for the simplest form of an analysis of covariance, and test at once the homogeneity of the regressions therein. We may set up the regression equation $H = a + bB$, where H represents the heart weight and B the total body weight. If b is actually less than the average percentage, .39%, then a must be significantly greater than 0 and vice versa for b greater than .39%. a will be significant if the residual variance is significantly reduced by the inclusion of a in the regression equation. The partition of the variation in heart weight after adjustment for body weight by covariance furnishes such a test. Thus for females, since $(1029.62)^2/265.13 = 3998.481$, we have a remainder of 66.229 g.^{2} for 46 degrees of freedom. Consequently, the significance of the observed difference between the regression value for heart weight and the average value can be tested by the following table.

TABLE 3

VARIATION IN HEART WEIGHT ADJUSTED FOR BODY WEIGHT

	Females			*Males*		
	d.f.	*(Heart)*^{2}	*M.S.*	*d.f.*	*(Heart)*^{2}	*M.S.*
From original totals	46	66.229	.....	96	233.669	.....
After correction for mean	45	60.787	1.351	95	230.259	2.424
Reduction due to mean	1	5.442	5.442	1	3.410	3.410

The reduction due to the mean for females is almost significant (the variance ratio is 4.03 while the 5% significance level is 4.06), indicating a strong probability that a is actually different from 0. Hence in

TABLE 4

ORIGINAL DATA ON LIVE BODY WEIGHT IN KILOGRAMS AND FRESH HEART WEIGHT IN GRAMS OF FEMALE AND MALE DOMESTIC CATS USED IN EXPERIMENTS ON DIGITALIS

Females

Body Weight	*Heart Weight*	*Body Weight*	*Heart Weight*	*Body Weight*	*Heart Weight*	*Body Weight*	*Heart Weight*
2.3	9.6	2.0	7.4	2.1	9.8	2.9	10.1
3.0	10.6	2.3	7.3	2.7	10.8	3.0	13.0
2.9	9.9	2.2	7.1	2.2	9.1	2.2	8.7
2.4	8.7	2.3	9.0	2.3	11.2	2.4	6.3
2.3	10.1	2.1	7.6	2.1	8.1	2.4	8.8
2.0	7.0	2.0	9.5	2.4	10.2	2.5	10.9
2.2	11.0	2.9	10.1	2.7	8.5	2.5	9.0
2.1	8.2	2.7	10.2	2.3	10.1	2.3	9.7
2.3	9.0	2.6	10.1	2.1	8.7	2.3	8.4
2.1	7.3	2.3	9.5	2.2	10.9	2.6	10.1
2.1	8.5	2.6	8.7	2.3	7.9	2.3	10.6
2.2	9.7	2.1	7.2	2.1	8.3	.....	.....

Males

Body Weight	*Heart Weight*	*Body Weight*	*Heart Weight*	*Body Weight*	*Heart Weight*	*Body Weight*	*Heart Weight*
2.9	9.4	2.5	12.7	3.5	15.7	3.5	17.2
2.4	9.3	3.5	15.6	2.8	13.3	3.8	16.8
2.2	7.2	2.4	9.1	2.2	9.1	2.2	8.5
2.9	11.3	2.2	7.6	2.5	7.9	3.3	15.4
2.5	8.8	3.4	12.8	2.4	7.9	2.7	9.8
3.1	9.9	2.6	8.3	3.9	14.4	3.2	11.9
3.0	13.3	3.4	11.2	3.1	12.5	2.9	10.6
2.5	12.7	2.6	9.4	3.7	11.0	3.6	13.3
3.4	14.4	2.7	8.0	3.0	12.4	2.7	12.5
3.0	10.0	3.3	14.9	3.2	13.5	2.9	11.8
2.6	10.5	2.2	10.7	3.3	14.1	3.6	15.0
2.5	8.6	3.2	13.6	3.0	12.7	2.8	10.2
2.8	10.0	2.2	9.6	2.9	10.1	2.5	11.0
3.1	12.1	3.5	11.7	3.0	10.4	2.6	11.5
3.0	13.8	2.5	9.3	2.4	7.9	3.9	20.5
2.7	12.0	3.2	12.3	3.8	14.8	3.0	12.2
2.8	12.0	3.2	13.0	2.0	6.5	2.6	9.4
2.1	10.1	2.7	9.6	3.1	11.5	2.7	9.0
3.3	11.5	2.6	7.7	2.8	9.1	2.5	8.8
3.4	12.2	2.7	9.6	2.3	9.6	2.2	9.6
2.8	13.5	2.0	6.5	3.0	11.6	3.1	13.0
2.7	10.4	3.1	14.3	2.2	7.9	3.3	12.0
3.2	11.6	2.4	7.3	3.4	12.4	2.7	11.1
3.0	10.6	3.6	14.8	3.5	12.9	3.6	11.8
.....	.....	.....	.....	.....	.....	2.8	11.4

females heart weight probably increases less than proportionately to body weight. In the case of males, where the heart weight appears to increase more than proportionately to the body weight, the differ-

ence is not significant with these data. The close agreement between the sexes in the average percentage of the body taken up by the heart seems to mask a real difference in the heart weight to be expected for a given body weight.

It may be noted that the estimated variance of heart weight for given body weight in males, 2.424 g.^{2}, is considerably greater than the value for females, 1.351 g.^{2} The greater residual variance for males possibly was related to their larger size. The heaviest female weighed 3.0 Kg. while nearly 40% of the males exceeded this weight.

Entered as second-class matter, May 25, 1945, at the post office at Washington, D. C., under the Act of March 3, 1879. *Biometrics* is published four times a year—in March, June, September and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N. W., Washington 6, D. C.

Membership dues in the American Statistical Association are $5.00 a year, of which $3.00 is for a year's subscription to the quarterly *Journal*, fifty cents is for a year's subscription to *The American Statistician*. Dues for Associate members of the Biometrics Section are $2.00 a year, of which $1.00 is for a year's subscription to *Biometrics*. Single copies of *Biometrics* are $1.00 each and annual subscriptions are $2.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N. W., Washington 6, D. C.

2×2 FACTORIAL EXPERIMENTS IN INCOMPLETE GROUPS FOR USE IN BIOLOGICAL ASSAYS

C. I. Bliss

Connecticut Agricultural Experiment Station and Yale University

Among the simpler experimental designs one of the most useful is the 2×2 factorial for testing two levels of two factors in all combinations. In developing an insecticidal treatment, for example, two concentrations of an oil emulsion may be applied with two different stickers. In a biological assay one drug may be compared with another at each of two comparable dosage levels. The two-dose design has been used widely in assays based upon a graded response and provides adequate estimates of relative potency when the form of the dosage-response curve is reasonably well known.

The two-dose assay owes some of its popularity to the greater ease of assembling homogeneous groups of four than of six or eight responses, which would be required for three or four dose assays. Even groups of four may be inconveniently large. When they consist of four successive reactions of the same animal separated by rest periods, the experiment is time-consuming and some individuals may die before it can be completed. When the groups are litters of the same sex the number with four males or four females may be limited. For these and other reasons groups of three or two may be preferred. Some experimenters [5, 6, 7] have used pairs and confounded one of the treatment effects with differences between pairs. This increases the error of the confounded effect. Frequently one would rather keep the same precision in all treatment comparisons and yet reduce the size of the group.

The experimental design known as balanced incomplete blocks or groups meets this need. As originally described by Yates [9], only the information within groups was utilized. The method was adapted to two-dose assays arranged in pairs by Bliss and Rose [2]. However, groups with different treatment combinations also provide estimates of the effect of treatment. More recently Yates [10] has described statistical methods for recovering inter-block information and has applied them to determining the adjusted mean response for each treatment and its error. In the 2×2 factorial design interest centers in three specific comparisons. Given this more limited objective, the calculation can be reduced to a convenient form.

Two restrictions in design. Frequently two restrictions are re-

quired in biological assays, such as occur in a Latin square. This is quite compatible with a balanced incomplete design. In a rabbit assay of insulin, for example, the responses of individual rabbits, the dates of testing and the doses of drug might form the rows, the columns and the letters of a 4×4 Latin square [1]. If the experiment were stopped after the third injection, the design would be an incomplete Latin square with the last column missing. Four rabbits would provide a set, but additional sets could be added to gain the necessary precision. In other cases the response may be modified by the number of times an animal has been used rather than by the environmental conditions on the day of testing, the columns then being the order of treatment, and the rows the individual animals. The 3×4 Latin rectangle, therefore, is a balanced incomplete design suitable for many biological assays.

The analysis of incomplete Latin squares as described by Yates [8] is based entirely on the information within rows, columns and letters. It is thus subject to the same limitations as the original incomplete block design. The estimated effects of treatment in the two cases are identical and the error differs only by the segregation of the differences between the three complete columns. For this reason the calculation of the incomplete Latin square as originally described by Yates will be replaced by that in his later paper.

The 2×2 factorial design in six balanced incomplete pairs can also be extended to segregate a second source of variation. Six pairs provide all possible combinations of the four doses in an assay. If order of treatment is a potential source of error, the minimal number can be duplicated and its order reversed. Since even the basic set of 12 pairs may not give sufficient replication, a second restriction on the design would seldom be burdensome.

Thus two restrictions in design are compatible with 2×2 factorial experiments arranged in incomplete groups of three or of two units. Both numerical examples are of this type. A somewhat more elaborate table is needed for the original data and the analysis of variance has one more row than in assays having only the restrictions represented by the incomplete triplets or pairs. In other respects the analysis is the same.

Experimental data. The method will be illustrated by two assays of parathyroid extract from the serum calcium of dogs [2]. The first example consisted originally of five 4×4 Latin squares in which each row represented the responses of a single dog and each column one

of four different days of treatment. The letters designated two doses of 0.125 and 0.205 cc. per kg. of a "standard" preparation and two doses in the same ratio of an "unknown." The data from the fourth test day have been omitted here, leaving the three observations on each of 20 dogs in Table 1. The rows in each Latin square have been rear-

TABLE 1

ASSAY OF PARATHYROID EXTRACT FROM 3 DETERMINATIONS ON EACH DOG OF THE SERUM CALCIUM IN MG. PER CENT (−10) 17 HOURS AFTER INJECTION; DOSAGE RATIO $U_2/U_1 = S_2/S_1 = 1.640$. DATA OF BLISS AND ROSE (2)

Set No.	Dog in group	Dose injected on			Serum calcium, y			
		3/15	3/24	4/5	3/16	3/25	4/6	Total, T_r
1	I	U_2	U_1	S_1	5.8	4.3	4.8	14.9
	II	U_1	U_2	S_2	4.7	5.4	4.8	14.9
	III	S_2	S_1	U_1	5.0	4.5	4.0	13.5
	IV	S_1	S_2	U_2	3.8	7.0	6.0	16.8
2	I	U_2	U_1	S_1	7.0	6.5	5.0	18.5
	II	U_1	U_2	S_2	5.1	5.0	5.8	15.9
	III	S_2	S_1	U_1	5.0	4.0	4.6	13.6
	IV	S_1	S_2	U_2	2.0	3.8	4.0	9.8
3	I	U_1	S_1	U_2	3.6	5.3	7.2	16.1
	II	U_2	S_2	U_1	4.4	3.8	4.4	12.6
	III	S_1	U_1	S_2	5.8	5.0	5.2	16.0
	IV	S_2	U_2	S_1	4.6	5.4	4.0	14.0
4	I	U_1	U_2	S_1	4.0	3.8	4.0	11.8
	II	U_2	U_1	S_2	6.2	4.0	3.0	13.2
	III	S_1	S_2	U_1	3.2	6.0	4.9	14.1
	IV	S_2	S_1	U_2	3.0	4.0	4.0	11.0
5	I	U_1	S_1	U_2	3.0	3.4	3.8	10.2
	II	U_2	S_2	U_1	5.8	6.0	5.0	16.8
	III	S_1	U_1	S_2	4.2	4.1	5.0	13.3
	IV	S_2	U_2	S_1	5.2	6.2	5.0	16.4
Total, T_c					91.4	97.5	94.5	283.4
Group Total, T_g		I 71.5		II 73.4	III 70.5		IV 68.0	Total, G 283.4

ranged so that the different combinations or groups of doses occur in a standard order. Since all dogs were treated at the same time, the 20 observations on each day could be handled as a unit in segregating the variation between days.

The second example required two injections in each dog separated by an interval of about two weeks. Twelve dogs comprised a basic set, the first six receiving the same treatment combinations as the sec-

ond six but in the reverse order (Table 6). Three such basic sets were completed, but unlike the Latin squares they were tested on different days with doses of 0.06 and 0.12 cc. of extract per kg. The dogs were more variable in size, from 6.4 to 16.4 kilograms. Except for proportioning the doses to the body weight, adjustments for differences in size have been omitted for the sake of simplicity.

ASSAYS IN BALANCED GROUPS OF THREE

Analysis of variance. This design is represented by the 3×4 Latin rectangle with f replications of the basic set of 12 responses. An experiment may be replicated at different times, but here we will consider simultaneous treatments. Each set is randomized independently by the procedure customary for Latin squares. We will adopt the convention that each row represents three responses by a single animal, each column one of the three different days of the test and each letter a different dose of drug. Two doses, S_1 and S_2, are of a standard preparation and the other two, U_1 and U_2, are equivalent doses of an unknown preparation, so far as its relative potency can be estimated in advance. The ratio of the two doses for the unknown must be the same as that for the standard. The designation of rows in any given case, of course, will depend upon the assay and often there may be no factor corresponding to columns. In all cases, however, four rows or groups are needed for each basic set.

It is convenient to arrange the rows of the basic set in a standard order for analysis. Each row will lack a different treatment and on this basis may be identified with one of four groups. The Roman numerals I to IV will designate groups without the treatments S_2, S_1, U_2 and U_1 respectively. We then need the total for each column or day (T_c), for each row or animal within each set (T_r), for each of the four groups over all sets (T_g), and for each dose separately within each group (T_d). The sums of each series of partial totals must check one another and equal the grand total (G).

An analysis of variance is computed primarily to estimate two error variances, one within groups and the other between groups. The equations for computing the sums of squares are shown in Table 2. Those in the first four lines are based upon the marginal totals in the table of original data. From the individual entries we obtain the total sum of squares in line 8 and from the grand total the correction for the general mean (C), which is usually computed first of all.

The above terms in the analysis of variance for our first example

TABLE 2

CALCULATION OF THE SUMS OF SQUARES FOR THE ANALYSIS OF VARIANCE OF EXPERIMENTS WITH TWO RESTRICTIONS WHEN ARRANGED IN f BASIC CONCURRENT SETS EACH OF FOUR TRIPLETS

Line No.	Variation due to	*D.F.*	Sum of squares	Mean square is estimate of
1	Columns (days)	2	$S(T_c^2)/4f - C$	
2	Individual rows (animals)	$4f-1$	$S(T_r^2)/3 - C$	
3	Groups I to IV	3	$S(T_g^2)/3f - C$	
4	Rows of same group	$4f-4$	Line 2 – Line 3	$\sigma_1^2 + 3\,\sigma_2^2$
5	Doses, ignoring groups	3	$S(V^2)/12f$	
6	Groups, eliminating dose	3	$S(W^2)/24f$	$\sigma_1^2 + 8\sigma_2^2/3$
7	Intra-row error	$8f-5$	Line 8 – Lines 1, 4, 5 and 6	σ_1^2
8	Total	$12f-1$	$S(y^2) - C$	
9	Correction for mean	1	$C = S^2(y)/12f$	
10	Inter-row error	$4f-1$	Lines 4 + 6	$\sigma_1^2 + \frac{4(3f-1)}{4f-1}\sigma_2^2$

have been computed directly from the data in Table 1. The first step was to determine from the grand total the correction for the mean as $C = 283.4^2/60 = 1338.593$. Substituting the partial totals in the formulae of Table 2 with $f = 5$, the sums of squares between days, between rows, between groups and the total followed easily. The results have been listed in Table 3. The sum of squares for groups I to IV ignoring dose was subtracted from that for rows to obtain the sum of squares between complete replicates or between dogs receiving the same combinations of treatments.

TABLE 3

ANALYSIS OF VARIANCE FOR PARATHYROID ASSAY IN TABLE 2

Variation due to	Degrees of freedom	Sum of squares	Mean square
Days (from T_c)	2	0.930	0.465
Rows (from T_r)	19	34.727	
Groups I to IV (from T_g)	3	1.011	
Dogs of same group (by difference)	16	33.716	2.1072
Doses, ignoring groups (from V)	3	10.946	
Groups, eliminating dose (from W)	3	1.941	0.6470
Intra-row error (by difference)	35	20.274	$0.5793 = s_1^2$
Total (from y)	59	67.807	
Correction for mean (from G)	1	1338.593	
Inter-row error	19	35.657	$1.8767 = A'$

The remaining two sums of squares are computed factorially. Factorial analysis may be characterized by factorial coefficients such as are given in the body of Table 4. The four treatments, U_1, U_2, S_1 and S_2 are shown at the top of the table and under each are three columns for the three groups (I through IV) in which it has been tested. There are thus 12 columns of coefficients. Below the coefficients is a row for the sums of the responses (T_d) from the f replicates of each group within each treatment. Of the 11 degrees of freedom among these partial totals, six are isolated individually by factorial analysis. The remaining five form part of the intra-group error which is computed as a difference.

The terms obtained factorially have been selected with two objectives in view, to complete the analysis of variance and to evaluate the assay. They are arranged in three pairs. The first line of each pair measures the crude effect of treatment, ignoring the subdivision into groups. The second line gives the net difference between groups associated with each specific treatment effect, after eliminating the effect of treatment. It is thus a component of the error. The treatments are the two direct effects and their interaction. Here they represent (a) the comparison of the standard with the unknown, (b) the combined slope of the dosage-response curves for both standard and unknown and (ab) their interaction or the divergence of these two curves from parallelism. The factorial coefficients in each row are multiplied by the corresponding treatment totals (T_d) in the last row of the table and the products summed. The sums of the products $S(xT_d)$ in the last column are designated as V and W respectively within each pair, with subscripts to indicate the treatment comparison.

The factorial coefficients (x) meet two requirements. In each row their sum is zero and in any two rows the products of the corresponding coefficients add up to zero. Hence they are mutually orthogonal and the variances computed from the sums of products are independent of each other. In completing the analysis of variance in Table 2, the three values of V are squared, summed and divided by $12f$ to obtain the sum of squares for doses ignoring groups. Similarly the three values of W are squared, summed and divided by $24f$ to obtain the sum of squares for groups eliminating dose. The values 12 and 24 are the sums of the squares of the factorial coefficients in a given row or $S(x^2)$. The seventh line in Table 2 is the sum of squares remaining after subtracting those in lines 1, 4, 5 and 6 from the total in line 8. It measures the variation within rows and columns. The

inter-group variation in the tenth line totals the sums of squares in lines 4 and 6.

Each treatment total (T_d) in Table 4 was obtained from Table 1 as the sum of the five observations for a given dose in one of three groups. These totals were multiplied in turn by the coefficients of each line in the table and the products summed to obtain the listed values for V and for W. From the V's in the first row of each treatment comparison—11.0, 23.0 and 2.6—we obtained the sum of squares (10.946) for doses ignoring groups and from the three W's in the second row for each term the sum of squares (1.941) for groups eliminating dose. Each represented three degrees of freedom. The analysis of variance was completed by computing the intra-group error as a difference and adding the two sums of squares for the inter-group variation. Each sum of squares was divided by the corresponding degrees of freedom to obtain the mean squares in Table 3.

The two components of experimental error. The analysis of variance in Table 2 provides estimates of two error terms, that within groups and that between groups. Our best estimate of the true error variance within groups (σ_1^2) is the mean square, s_1^2, in line 7. The inter-group errors in lines 4 and 6 have been computed from two sources. The mean square in line 4 is based upon complete replicates and by component analysis [4] it is an estimate of $\sigma_1^2 + 3\sigma_2^2$, where σ_2^2 is the population variance between groups. The mean square in line 6 is modified by the efficiency (E) of the experiment which here is equal to 8/9 as given by Yates [10]. Hence it is an estimate of $\sigma_1^2 + E(3\sigma_2^2)$.

The sums of squares in lines 4 and 6 are added in the last row of Table 2, giving a mean square designated as A'. Because the components from lines 4 and 6 have been weighted by their degrees of freedom, A' is the composite in the last column of the table. Replacing the population variances σ_1^2 and σ_2^2 by their estimates, s_1^2 and s_2^2, and solving for s_2^2, the net variance between rows can be computed from the mean square in line 10 as

$$s_2^2 = \frac{4f-1}{12f-4}(A' - s_1^2). \qquad \text{.....(1)}$$

In experiments where the basic sets of 12 responses are not carried out together, the differences between sets should be segregated from the sum of squares in line 4. This would reduce the degrees of freedom used as its weight in estimating s_2^2 and in consequence would modify Equation 1 and all terms depending upon it.

TABLE 4

FACTORIAL ANALYSIS OF THE EFFECTS OF TREATMENT AND ADJUSTMENT FOR DIFFERENCES BETWEEN GROUPS OF 3, DATA FROM TABLE 2

Term	Factorial coefficients (x) for treatments in each group												Sum of products $S(xT_d)$
	U_1			U_2			S_1			S_2			
	I	II	III	I	II	IV	I	III	IV	II	III	IV	
(a) S vs. U	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	$11.0 = V_a$
	−1	−1	2	−1	−1	2	−2	1	1	−2	1	1	$-4.1 = W_a$
(b) Slope	−1	−1	−1	1	1	1	−1	−1	−1	1	1	1	$23.0 = V_b$
	1	−2	1	2	−1	−1	1	1	−2	−1	2	−1	$12.4 = W_b$
(ab) Parallel	−1	−1	−1	1	1	1	1	1	1	−1	−1	−1	$2.6 = V_{ab}$
	−2	1	1	−1	2	−1	−1	2	−1	1	1	−2	$7.9 = W_{ab}$
Dose total, T_d	21.4	23.2	22.6	27.6	26.8	25.6	22.5	21.7	18.8	23.4	26.2	23.6	

TABLE 5

FACTORIAL ESTIMATION OF THE TREATMENT EFFECT S VS. U FROM INTRA-GROUP AND FROM INTER-GROUP DIFFERENCES COMPARED WITH THE EQUIVALENT COEFFICIENTS FROM TABLE 4. NUMERICAL VALUES OBTAINED WITH THE VALUES OF T_d IN THE LAST ROW OF TABLE 4

Term	Factorial coefficients (x) for treatments in each group												Divisor $fS(x^2)$	Sum of products $S(xT_d)$	Variance $\frac{S^2(xT_d)}{fS(x^2)}$
	I			II			III			IV					
	U_1	U_2	S_1	U_1	U_2	S_2	U_1	S_1	S_2	U_2	S_1	S_2			
Intra-block	1	1	−2	1	1	−2	2	−1	−1	2	−1	−1	$24f$	$T_i = 13.3$	1.4741
Inter-block	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	$12f$	$T_j = 6.4$	0.6827
Crude effect	1	1	−1	1	1	−1	1	−1	−1	1	−1	−1	$12f$	$V = 11.0$	2.0167
Adjustment	−1	−1	−2	−1	−1	−2	2	1	1	2	1	1	$24f$	$W = -4.1$	0.1401

If the arrangement in groups does not segregate a disturbing source of variation, the mean square between groups A' should be equal to or smaller than that within groups (s_1^2). In this case treatment effects could be computed from the totals for each dose without considering differences between groups and the error variance would be s_1^2. Usually, however, A' is larger than s_1^2. Treatments are then adjusted for differences between groups and are subject to a somewhat larger error than if this adjustment were unnecessary.

Numerical values have been obtained with the above procedure from the analysis of variance in Table 3. The mean square for intra-group error was $s_1^2 = 0.5793$ with 35 degrees of freedom. The sums of squares for the variation among dogs were added and divided by $16 + 3 = 19$ to obtain the mean square between groups of $A' = 1.8767$, which exceeded s_1^2 significantly. Substituting in Equation 1 the estimated net variance between groups was $s_2^2 = \frac{19}{56}(1.8767 - 0.5793) = 0.4402$. The net variance between dogs was nearly as large as that for different readings on the same dog.

Determination of adjusted effects of treatment. Two independent estimates of each treatment effect are obtainable from an experiment in incomplete balanced groups. One is based upon comparisons within groups and the other upon comparisons between groups. The most satisfactory measure of a treatment comparison is the weighted mean of the intra-group and inter-group estimates. The same result can be computed alternatively and more conveniently from the values of V and W computed with the coefficients in any pair of rows in Table 4. These provide respectively a crude estimate of the treatment effect and a factor for its adjustment. The adjustment is first multiplied by a coefficient μ, which is computed from the error variances, s_1^2 and s_2^2. After a crude effect V is adjusted with μ and W, the result must agree with the weighted mean from the intra-group and inter-group comparisons.

The factorial coefficients in the first two lines of Table 5 lead respectively to the intra-group and inter-group treatment comparisons. Both represent the comparison of standard vs. unknown; the others follow a similar pattern. The intra-group estimate (T_i) is equivalent to that developed initially for the analysis of incomplete blocks [9]. Its variance may be computed as $V(T_i) = fS(x^2)s_1^2 = 24fs_1^2$. In units of a single *response*, the intra-group estimate is

$$t_i = T_i/fS\,|x| = T_i/16f,$$

which has a variance of

$$V(t_i) = \frac{fS(x^2)}{f^2S^2|x|}\, s_1^2 = \frac{3s_1^2}{32f} = \frac{1}{w_i}.$$

The inter-group estimate, T_j, in the second line has an observed variance of $fS(x^2)(s_1^2 + 3s_2^2) = 12f(s_1^2 + 3s_2^2)$. In units of a single *group* of three responses, the inter-group estimate may be written as

$$t_j = T_j/4f$$

and its variance as

$$V(t_j) = 12f(s_1^2 + 3s_2^2)/16f^2 = 3(s_1^2 + 3s_2^2)/4f = \frac{1}{w_j}.$$

These two estimates are averaged, weighting each by the reciprocal of its variance, to obtain the adjusted treatment effect

$$t_a = \frac{w_i t_i + w_j t_j}{w_i + w_j}. \qquad \text{.....(2)}$$

The terms t_i and t_j in Equation 2 may be expressed in units of V and W. The coefficients for the crude effect V and for the adjustment W have been copied from Table 4 to Table 5. The coefficients in the last two rows of Table 5 measure identically the same treatment effect as those in the first two rows. Thus the products of the corresponding coefficients in the first pair differ only in sign from those in the second. A second check is to compare the variances in the last column of the table, which were computed with the T_d's of our numerical example. The first two variances totalled $1.4741 + 0.6827 = 2.1568$ and the next two $2.0167 + 0.1401 = 2.1568$. The factors for expressing T_i and T_j in terms of V and W can be obtained by solving two pairs of simultaneous equations based upon the coefficients for U_2 and for S_1 in Group I. From the solution of these equations we find that

$$T_i = (4V + W)/3 \text{ and } T_j = (V - 2W)/3,$$

which can be checked with the numerical values in Table 5. Substituting in Equation 2 and simplifying we have

$$t_a = \frac{V}{12f} + \frac{w_i - 8w_j}{w_i + w_j} \cdot \frac{W}{48f} \qquad \text{.....(3)}$$

Each adjusted comparison t_a may be expressed conveniently in the same units as V by multiplying it by $12f$. For the comparison of standard and unknown

$$T_a = V_a + 2\mu W_a, \qquad \text{.....(4)}$$

where $\mu = \dfrac{w_i - 8w_j}{8(w_i + w_j)}$ and the subscripts a, b or ab refer to a given factorial effect. In terms of the estimated variances within and between groups, $\mu = s_2^2/(3s_1^2 + 8s_2^2)$. Substituting from Equation 1

$$\mu = \frac{(4f-1)(A' - s_1^2)}{8(4f-1)A' + 4(f-1)s_1^2}. \qquad \text{.....(5)}$$

The three adjusted treatment comparisons may be determined for our numerical example. Substituting the values for s_1^2 and A' from Table 3 with $f = 5$, $\mu = \dfrac{19 \times 1.2974}{152 \times 1.8767 + 16 \times 0.5793} = 0.08370$. This is used in Equation 4 with the sums of products computed factorially in Table 4 to obtain the three adjusted effects of treatment of $T_a = 11.0 - 2 \times 0.08370 \times 4.1 = 10.314$, $T_b = 25.076$ and $T_{ab} = 3.922$.

The error of the factorial comparisons. When the variation between groups (A') is larger than that within groups (s_1^2), the crude treatment comparisons are adjusted with the information between groups and the error for evaluating the adjusted comparison exceeds s_1^2. In units of a single response the adjusted factorial comparison, t_a, is the weighted mean of two components. Its variance is equal to the reciprocal of the sum of the weights or

$$V(t_a) = \frac{1}{w_i + w_j} = \frac{s_1^2}{4f} \cdot \frac{s_1^2 + 3s_2^2}{3s_1^2 + 8s_2^2}.$$

Since $1 + \mu = \dfrac{3(s_1^2 + 3s_2^2)}{3s_1^2 + 8s_2^2}$, this may be reduced to

$$V(t_a) = \frac{s_1^2}{12f}(1 + \mu).$$

In practice, however, we are more interested in $T_a = 12ft_a$ which has a standard error of

$$s = \sqrt{12f(1 + \mu)s_1^2}. \qquad \text{.....(6)}$$

Each of the adjusted factorial comparisons, T_a, T_b, and T_{ab}, has the same error. For testing their significance, a "just significant difference" may be computed as

$$c = ts, \qquad \text{.....(7)}$$

where t is taken from a table of Student's t with the degrees of freedom in s_1^2 and the desired level of significance, usually at $P = 0.05$.

Substituting the numerical values from our example, the standard error of each adjusted treatment effect was computed by Equation 6

as $s = \sqrt{60 \times 1.0837 \times 0.5793} = 6.137$. With 35 degrees of freedom the tabular value of t at $P = 0.05$ is 2.032, giving a "just significant difference" of $c = 2.032 \times 6.137 = 12.470$ for evaluating the three adjusted factorial effects. In comparison with these criteria, the two dosage-effect curves were parallel (T_{ab}) well within the sampling error, the combined slope (T_b) was highly significant and the potency of the unknown did not differ significantly from that of the standard (T_a).

THE BALANCED PAIR DESIGN

Analysis of variance. Groups of two may be required in some experiments by the nature of the test material or by the exigencies of the assay. When each treatment is paired once with every other, six groups will cover all possible combinations of four treatments. However, if the two responses in a pair represent successive reactions of the same animal, the six pairs of a series should be matched by a second series in which the treatments are given in the reverse order. This leads to the experimental design shown in the left side of Table 6.

TABLE 6

ASSAY OF PARATHYROID EXTRACT FROM 2 DETERMINATIONS ON EACH DOG OF THE SERUM CALCIUM IN MG. PER CENT (−10) 17 HOURS AFTER INJECTION; DOSAGE RATIO $U_2/U_1 = S_2/S_1 = 2.00$. DATA OF BLISS AND ROSE (2)

Dog with group	Treatment 1st	Treatment 2nd	Serum calcium, y								
			First set			Second set			Third set		
			9/20	10/4	T_r	9/21	10/5	T_r	9/22	10/6	T_r
I	U_1	U_2	2.0	4.4	6.4	4.0	5.9	9.9	4.2	4.6	8.8
II	U_1	S_1	3.4	2.1	5.5	3.8	3.8	7.6	2.6	3.4	6.0
III	U_1	S_2	2.2	2.2	4.4	2.8	4.0	6.8	3.4	4.8	8.2
IV	U_2	S_1	3.4	2.1	5.5	4.8	4.0	8.8	4.1	3.0	7.1
V	U_2	S_2	2.4	4.2	6.6	3.4	3.6	7.0	3.0	4.0	7.0
VI	S_1	S_2	2.0	2.4	4.4	1.8	2.2	4.0	3.7	5.2	8.9
I	U_2	U_1	3.2	2.8	6.0	3.6	2.0	5.6	4.2	3.8	8.0
II	S_1	U_1	1.0	2.5	3.5	2.9	2.9	5.8	3.0	3.3	6.3
III	S_2	U_1	4.2	2.2	6.4	2.6	3.8	6.4	5.4	3.0	8.4
IV	S_1	U_2	1.6	2.4	4.0	1.6	4.0	5.6	4.0	8.8	12.8
V	S_2	U_2	4.2	4.8	9.0	3.0	5.2	8.2	4.4	4.1	8.5
VI	S_2	S_1	3.4	1.6	5.0	5.2	4.0	9.2	4.6	3.2	7.8
Total, T_c and T_s			33.0	33.7	66.7	39.5	45.4	84.9	46.6	51.2	97.8
Group			I	II	III	IV	V	VI	Total, G		
Total, T_g			44.7	34.7	40.6	43.8	46.3	39.3	249.4		

Even with 12 pairs in the basic set, f replicates may be required. While an assay in symmetrical pairs can be computed entirely from

the differences within groups [2], the pair totals may furnish an appreciable amount of information. A more complete analysis differs but little from that for groups of three. In the following discussion we will assume that the basic set consists of 12 pairs and that there are f replicates of the basic set.

The first stage in the calculation is to determine the column or day totals (T_c), the row or pair totals (T_r) and the totals for each of the six groups, (T_g). When the experiments are concurrent, the first four rows of the analysis of variance follow a pattern similar to that in Table 2. When the f basic sets of 12 pairs are not treated together, the sum of squares between complete replicates must be segregated as an additional term. It is computed from the f totals (T_s) of the differently-timed sets as shown in the first line of Table 7. Within basic sets the f differences between column totals (T_c) are independent of either treatments or rows. When the replicates are not concurrent, these totals are squared separately for each set, leading to a sum of squares between columns with f degrees of freedom (line 2, Table 7).

TABLE 7

CALCULATION OF THE SUMS OF SQUARES FOR THE ANALYSIS OF VARIANCE OF EXPERIMENTS WITH TWO RESTRICTIONS WHEN ARRANGED IN f BASIC SETS OF 12 PAIRS CARRIED OUT AT DIFFERENT TIMES

Line No.	Variation due to	Degrees of freedom	Sum of squares	Mean square is estimate of
1	Complete sets or replicates	$f-1$	$S(T_s^2)/24-C$	
2	Columns within sets (days)	f	$S(T_c^2)/12-C-$ Line 1	
3	Individual rows (animals)	$11f$ 5	$S(T_r^2)/2-C-$ Line 1	
4	Groups I to VI	$11f-5$	$S(T_g^2)/4f-C$	
5	Rows of same group	3	Line 3 – Line 4	$\sigma_1^2+2\sigma_2^2$
6	Doses, ignoring groups	3	$S(V^2)/24f$	
7	Groups, eliminating dose	2	$S[W^2/2fS(x^2)]$	$\sigma_1^2+4\sigma_2^2/3$ $\sigma_1^2+2\sigma_3^2$
8	Intra-row error	$11f-3$	Line 9 – Lines 1, 2, 5, 6, 7	σ_1^2
9	Total	$24f-1$	$S(y^2)-C$	
10	Correction for mean	1	$C=S^2(y)/24f$	
11	Inter-row error	$11f$	Lines 5 + 7	$\sigma_1^2+\frac{2(11f-1)}{11f}\sigma_2^2$

Since it consisted of three basic sets tested at different times, the parathyroid assay in balanced pairs of Table 6 illustrates this second

type. The original values for each dog were summed to obtain the total for each pair (T_r), for each day (T_c), for each set (T_s) and, over all three replicates, for each of the six groups (T_g). The sums of squares based upon these totals were computed as shown in the work form of Table 7 with the results given in Table 8.

TABLE 8

ANALYSIS OF VARIANCE FOR PARATHYROID ASSAY IN TABLE 6

Variation due to	Degrees of freedom	Sum of squares	Mean square
Complete sets (from T_s)	2	20.345	10.1725
Days within sets (from T_c)	3	2.353	0.7843
Rows within sets (from T_r)	33	45.301	
Groups I to VI (from T_g)	5	7.536	
Dogs of same group (by difference)	28	37.765	1.3488
Doses, ignoring groups (from V)	3	27.183	
Groups, eliminating dose (from W)	5	1.317	0.2634
Intra-row error (by difference)	30	18.403	$0.6134 = s_1^2$
Total (from y)	71	107.366	
Correction for mean (from G)	1	863.894	
Inter-row error	33	39.082	$1.1843 = A'$

The sums of squares associated with the different treatments in a paired experiment are computed factorially with the coefficients in Table 9. The responses to each dose are summed separately for each of the six treatment combinations in groups I to VI, including the inverted pairs, and the totals (T_d) entered beneath the columns of factorial coefficients. The sums of squares in lines 6 and 7 of the analysis of Table 7 are computed with these coefficients. That in line 6 is determined from the values of V and represents the crude effect of treatments or the differences in dose ignoring groups. As in the preceding case, the coefficients (x) are assigned to the doses, U_1, U_2, S_1 and S_2 without regard to the pairs in which they occur. The three sums of products (V) are squared, summed and divided by $2fS(x^2) = 24f$ to obtain the sum of squares for the analysis of variance.

The sum of squares for groups eliminating dose has 5 degrees of freedom. Three of the five are associated with the three treatment effects; their coefficients have been obtained from the equations for W_s as given by Yates [10]. The variation represented by the remaining two degrees of freedom is independent of that associated with treatment: their coefficients are shown in the last two rows of Table 9. The sum of products from each of these five rows (W_a to W_e) is squared,

TABLE 9

FACTORIAL ANALYSIS OF THE EFFECTS OF TREATMENT AND ADJUSTMENT FOR DIFFERENCES BETWEEN GROUPS OF 2, DATA FROM TABLE 7

Term	Factorial coefficients (x) for treatments in each group												Sum of products $S(xT_d)$
	U_1			U_2			S_1			S_2			
	I	II	III	I	IV	V	II	IV	VI	III	V	VI	
(*a*) S vs. T	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	$12.6 = V_a$
	−2	1	1	−2	1	1	−1	−1	2	−1	−1	2	$-3.6 = W_a$
(*b*) Slope	−1	−1	−1	1	1	1	−1	−1	−1	1	1	1	$42.4 = V_b$
	−1	2	−1	1	1	−2	2	−1	−1	1	−2	1	$7.6 = W_b$
(*ab*) Parallel	−1	−1	−1	1	1	1	1	1	1	−1	−1	−1	$0.8 = V_{ab}$
	−1	−1	2	1	−2	1	1	−2	1	2	−1	−1	$-8.8 = W_{ab}$
Inter-dog error	1	−1	0	1	0	−1	−1	0	1	0	−1	1	$3.0 = W_d$
	1	1	−2	1	−2	1	1	−2	1	−2	1	1	$-3.8 = W_e$
Dose total, T_d	18.8	18.5	17.4	25.9	27.5	22.9	16.2	16.3	16.3	23.2	23.4	23.0	

divided by $2fS(x^2)$ and the resulting variances added to obtain the sum of squares for groups eliminating dose. The remaining sum of squares, that for intra-pair error, is then obtained by difference from the total sum of squares, completing the analysis of variance.

The mean square for intra-row error in Table 7 is an estimate (s_1^2) of the population variance σ_1^2. The two sums of squares which are totalled to obtain the mean square for inter-row error (A') in the last line of Table 7 represent three components. The components from line 5 with $(11f-5)$ degrees of freedom and from W_d and W_e (Table 9) in line 7 with 2 degrees of freedom are based upon true replicates or their equivalent and estimate $\sigma_1^2+2\sigma_2^2$. The third component, computed from W_a, W_b and W_{ab}, is associated with the three treatment comparisons and hence is an estimate of $\sigma_1^2+E(2\sigma_2^2)$ where $E=4/6$. The mean square, (A'), based upon the three components is a composite of the two error variances as shown in the last column and row of Table 7. Substituting the estimate s_1^2 and s_2^2 for σ_1^2 and σ_2^2 and solving for s_2^2, the net error variance between rows is

$$s_2^2=\frac{11f}{2(11f-1)}(A'-s_1^2). \qquad \text{.....(8)}$$

The dose totals (T_d) from our numerical example in the last line of Table 9 were multiplied by the factorial coefficients in each row and the products summed to determine the required values of V and of W. The sum of squares for doses ignoring groups was calculated from the V's as $(12.6^2+42.4^2+0.8^2)/72=27.183$; that for groups eliminating dose was computed from the W's in the remaining rows as $(3.6^2+7.6^2+8.8^2+3.8^2)/144+3.0^2/48=1.317$. The intra-group error with 30 degrees of freedom was obtained as a difference, leading to $s_1^2=0.6134$. Finally, the two sums of squares for the variation between dogs were added to obtain the inter-group mean square in the last row of Table 8, as $A'=1.1843$ with 33 degrees of freedom. The mean squares s_1^2 and A' differed significantly although less than in the preceding experiment. The error variance between rows was estimated by Equation 8 as $s_2^2=\frac{33}{64}(1.1843-0.6134)=0.2944$.

Adjusted factorial effects and their error. The adjusted treatment effects from an experiment in balanced pairs resemble those for groups of three. Two independent estimates are available from the original data, one based upon intra-group and the other upon inter-group comparisons. The first (t_i) is expressed in units of a single response and

the second (t_j) in terms of a single pair of responses. By a process similar to that in Table 5, these estimates can be computed from the sums of products in Table 9 as

$$t_i = (2V + W)/48f \text{ and } t_j = (V - W)/12f.$$

In determining the adjusted treatment effect t_a by Equation 2, each component is weighted by the reciprocal of its variance or by

$$w_i = 16f/s_1^2 \text{ and } w_j = 2f/(s_1^2 + 2s_2^2).$$

The adjusted effects t_a are expressed more conveniently in the same units as V by the equation

$$T_a = 24ft_a = V + \frac{s_1^2 V + 2(s_1^2 + 4s_2^2)W}{9s_1^2 + 16s_2^2}. \quad \text{.....(9)}$$

When the f basic sets are not concurrent, s_2^2 may be replaced by its value in Equation 8 and the adjusted factorial effect computed as

$$T_a = V + \mu\{(1-k)V + 2(1+k)W\}, \quad \text{.....(10)}$$

where

$$\mu = \frac{s_1^2 + 2s_2^2}{9s_1^2 + 16s_2^2} = \frac{11fA' - s_1^2}{88fA' + (11f-9)s_1^2} \quad \text{.....(11)}$$

and

$$k = \frac{2s_2^2}{s_1^2 + 2s_2^2} = \frac{11f(A' - s_1^2)}{11fA' - s_1^2}. \quad \text{.....(12)}$$

The standard error of T_a can be computed readily as

$$s = \sqrt{\frac{(24f)^2}{w_i + w_j}} = \sqrt{288f\mu s_1^2} \quad \text{.....(13)}$$

which leads to c by Equation 7.

The above equations apply directly to the numerical example in Table 6. The coefficients μ and k were computed from A' and s_1^2 in Table 8 with $f = 3$, giving $\mu = \frac{38.4686}{327.3768} = 0.11751$ and $k = \frac{18.8397}{38.4686} = 0.48974$. The adjusted treatment comparisons could be determined by substituting the values for V and W of Table 9 in Equation 10 or

$$T_a = V_a + 0.1175\ (0.5103V_a + 2.9795W_a).$$

Then the adjusted values were $T_a = 12.095$, $T_b = 47.603$ and $T_{ab} = -2.233$. By equation 13 each adjusted total had a standard error of $s = \sqrt{864 \times 0.11751 \times 0.6134} = 7.892$. With 30 degrees of freedom in s_1^2, the tabular value of t at $P = 0.05$ was used in Equation 7 to obtain a "just significant difference" of $c = 2.042 \times 7.892 = 16.115$. When the

adjusted values were compared with these criteria, the experiment in balanced pairs led to the same conclusion as that in groups of three.

THE ESTIMATION OF POTENCY

The validity of a given assay may be judged from the relation between the adjusted treatment effects and their error. The value for T_b or slope should be highly significant and hence much larger than c. T_{ab}, on the other hand, should be less than c and preferably less than s. This would indicate that the dosage-response curves for standard and unknown were parallel within the experimental error. A value of T_{ab} larger than c and of T_a equal to or less than s would suggest a qualitative difference between the standard and the unknown such that their relative potencies would vary with the level of response. However, if both T_a and T_{ab} were statistically significant, one might conclude more probably that the doses had exceeded the linear range of the dosage-response curve. In this case the experiment should be repeated with a revised assumption as to the potency of the unknown. As has already been noted, both examples met these requirements for a valid assay.

The relative potency of the unknown may now be computed in the units which were assumed to be equal when setting up the assay. In terms of logarithms it is

$$M' = iT_a/T_b, \qquad \text{.....(14)}$$

where T_a and T_b are adjusted treatment effects and i is the interval between successive log-doses of drug. The proportionate relative potency or the anti-logarithm of M' provides the basis for adjusting the concentration of an unknown preparation to equivalence with the standard.

The standard error of M', of use in estimating approximate confidence limits, may be computed as

$$s_M = \frac{i\,s\sqrt{T_a^2 + T_b^2}}{T_b^2}. \qquad \text{.....(15)}$$

Approximate confidence limits are equal to

$$X_L' = M' \pm ts_M, \qquad \text{.....(16)}$$

where t is that used in computing c by Equation 7. With the small numbers in most assays, one would prefer (3) the exact confidence limits

$$X_L = \frac{i(T_aT_b \pm c\sqrt{T_a^2 + T_b^2 - c^2})}{T_b^2 - c^2}, \qquad \text{.....(17)}$$

where all terms have their prior significance. The anti-logarithms of these limits give the range in the relative potency of the unknown at the confidence level selected in computing c by Equation 7.

The relative potency of the unknown parathyroid extracts in our numerical examples has been computed with Equation 14. The experiment in groups of three used a dosage-ratio of 1.64, so that $i = \log(1.64) = 0.2148$. The log-potency was determined as $M' = 0.2148 \times 10.314/25.076 = 0.0883$. The best estimate of the potency of the unknown, relative to the standard, was the antilog of $2 + M'$ or 122.6 per cent. The equivalent values for the assay in balanced pairs with a greater dosage-ratio of $i = \log(2) = 0.3010$ were $M' = 0.0765$ and relative potency = 119.3 per cent. Hence the two unknowns agreed closely in potency.

The approximate and exact confidence limits have been compared at $P = 0.05$, using Equations 16 and 17 respectively. In percentage terms the approximate limits for the first example were 94 and 160 per cent and the exact limits 96 and 180 per cent. In comparison with the exact values the approximate values for the first example underestimated the limits of potency and the range. The two estimates agreed more closely in the second example, largely because the greater dosage interval increased the reliability of the slope. The approximate and exact confidence limits were 94–152 and 94–158 per cent respectively.

Summary. Factorial assays with two doses of the standard drug and two doses of the unknown have proved experimentally practicable and easy to analyze when based upon a graded response. In many cases potential sources of variation can be segregated by means of randomized block or Latin square designs. When homogeneous groups of four are less practicable than those of two or three, the balanced incomplete group may be substituted. The design and analysis of factorial experiments in balanced pairs and triplets are discussed and illustrated by two parathyroid assays based upon the serum calcium in dogs.

REFERENCES

1. Bliss, C. I. and H. P. Marks. "The Biological Assay of Insulin. II. The Estimation of Drug Potency from a Graded Response," *Quarterly Journal of Pharmacy and Pharmacology* 12 : 182–205, 1939.
2. Bliss, C. I. and C. L. Rose, "The Assay of Parathyroid Extract from the Serum Calcium of Dogs," *American Journal of Hygiene* 31 : 79–98, 1940.
3. Bliss, C. I., "Confidence Limits for Biological Assays," *Biometrics Bulletin* 1 : 57–65, 1945.
4. Crump, S. Lee, "The Estimation of Variance Components in Analysis of Variance," *Biometrics Bulletin* 2 : 7–11, 1946.
5. Gridgeman, N. T., "Special Designs for Vitamin D Assays," *Quarterly Journal of Pharmacy and Pharmacology* 18 : 15–23, 1945.

6. Price, W. C., "Accuracy of the Local-Lesion Method for Measuring Virus Activity, IV. Southern Bean Mosaic Virus," *American Journal of Botany* 32 :613–19, 1945.
7. Smith, K. W., H. P. Marks, E. C. Fieller and W. A. Broom, "An Extended Cross-Over Design and its Use in Insulin Assay," *Quarterly Journal of Pharmacy and Pharmacology* 17 : 108–123, 1944.
8. Yates, F., "Incomplete Latin Squares," *Journal of Agricultural Science* 26 : 301–315, 1936.
9. Yates, F., "Incomplete Randomized Blocks," *Annals of Eugenics* 7 : 121–140, 1936.
10. Yates, F., "The Recovery of Inter-Block Information in Balanced Incomplete Block Designs," *Annals of Eugenics* 10 : 317–325, 1940.

Officers of the American Statistical Association: President, Willard L. Thorp; Directors, Isador Lubin, Lowell J. Reed, Walter A. Shewhart, Samuel A. Stouffer, Helen M. Walker, Samuel S. Wilks; Vice-Presidents, Chester I. Bliss, Philip M. Hauser, Stacy May, Jacob Marschak, Jerzy Neyman, Frank W. Notestein, George W. Snedecor, Aryness Joy Wickens; Secretary-Treasurer, Lester S. Kellogg.

Officers of the Biometrics Section: Chairman, D. B. DeLury; Secretary, H. W. Norton; Section Committee, Geoffrey Beall, E. J. DeBeer, D. B. DeLury, D. J. Finney, H. W. Norton and J. W. Tukey.

Editorial Committee for *Biometrics:* Chairman, Gertrude M. Cox; Members, R. L. Anderson, C. I. Bliss, W. G. Cochran, Churchill Eisenhart, H. W. Norton, G. W. Snedecor and C. P. Winsor; collaborating editors, W. J. Dann, D. J. Finney, G. E. Dickerson, H. O. Halvorson, C. M. Mottley, J. G. Osborne.

Material for *Biometrics* should be addressed to the Chairman of the Editorial Committee, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to "Queries," Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.

STATISTICAL TREATMENT OF COUNTS OF TRICHOSTRONGYLID EGGS*

L. OTIS EMIK**
University of California, College of Agriculture, Davis

Counts of nematodes' eggs from fecal samples of the host have been used as a measure of the infestation level of the host although a multiplicity of factors may affect the egg worm ratio. While techniques have been improved for securing reliable counts from samples, statistical methods for analyzing these counts have seldom been used to full advantage. This is also true of other measures of infestation level.

The ultimate aims of making estimates of infestation levels and statistical analyses of them vary with the experiment, but statistically speaking, the purpose is essentially the same, i.e., to show the presence or absence of significant differences between individuals, groups of individuals, or time intervals in the same individual or group of individuals. The methods of analysis of variance could be applied to these types of data, in some cases without modification of experimental design. Gregory, Miller and Stewart [5], and Stewart, Miller and Douglas [7], tested differences between mean egg counts from sheep, breeds of sheep, different sires, and different ages by the individual pairing method, Snedecor [6] Section 2.13. Experimental design indicated the propriety of the method, but it was cumbersome and did not give the information afforded by analysis of variance. Ackert, Eisenbrandt, Wilmoth, Glading and Pratt [1], compared the resistance of five breeds of chickens to the nematode *Ascaridia lineata* by differences between means of number of worms and length of worms recovered at autopsy. Their data were suitable for analysis of variance and also covariance. More commonly, data on infestation levels have not been derived from experimental designs which would permit such analysis, but the experimental design usually could be altered to allow such analysis.

The data presented here are typical, in most respects, of those which have been used for determining relative susceptibility of sheep to intestinal nematodes. This report presents a method of transformation and an analysis of variance appropriate for the data.

Twelve lambs were used to obtain these data. They ranged from 157 days of age to 180 days of age, with one 126 days old, at the end

* Data taken from doctor's thesis.
** At present with the U. S. Sheep Experiment Station and Western Sheep Breeding Laboratory, Dubois, Idaho.

of the experiment. Nine were Hampshires from two relatively new sires in the flock. The last three represented the Shropshire, Rambouillet and Merino breeds.

TABLE 1

ACTUAL COUNTS OF PARASITES' EGGS. MEANS FOR LAMBS, CORRECTED OR TRANSFORMED

Lamb Number	Original counts	Mean square root	Corrected for moisture	Mean square root of corrected
2510	8.3	2.73	12.9	3.36
2563	5.5	2.22	7.6	2.62
2566	6.1	2.27	11.1	3.08
2569	13.9	3.61	19.8	4.33
2571	15.8	3.82	19.7	4.31
2574	17.9	4.16	27.5	5.11
2578	11.5	3.33	17.1	4.05
2580	16.4	3.91	28.8	5.20
2591	12.8	3.49	16.6	3.97
2622	6.4	2.46	7.5	2.67
2653	10.7	3.19	13.5	3.58
BT76	8.9	2.89	10.6	3.12
Mean	11.2	3.18	16.1	3.78

These lambs were running with their mothers in the University flock at the beginning of the experiment. All lambs were weaned and placed in one flock a few days later. Throughout the experiment the lambs grazed on Ladino clover during the evening and night, being corralled during the heat of the day.

Counts of nematodes' eggs were made over a period of 20 consecutive days. Fecal samples were collected each morning about 10 o'clock. The samples were prepared and counted according to the Stewart, Miller and Douglas [7] modification of the Caldwell method. Two counts were made on each sample. In making these counts a dilution technique was employed. One gram of feces was diluted to 10 ml. and two 0.1-ml. samples were completely counted. This is similar in principle to a hemacytometer count of blood cells, or the dilution method of counting bacteria. It is well established that such counts approach a Poisson distribution.

Since these counts represented 12 heterogeneous individuals with considerable daily variation, one appropriate test of the fit of the basic data to a Poisson distribution was available, Fisher [4], p. 61. The two counts made on each sample were separate estimates of the same population of eggs, and a test of their variance would test agreement

with a Poisson distribution. Chi-square tests of the fit of individual counts to sample means should be distributed according to the probability levels in the table of Chi-square, in this case with one degree of freedom. The Chi-square test, using the original uncorrected counts, Table 2, gave 6.25 with 7 degrees of freedom, a probability above 50 percent. This test indicated that the data were distributed in a Poisson fashion. Further, the variances of the means were highly significantly correlated to the means.

TABLE 2

TEST OF FIT OF ORIGINAL UNCORRECTED COUNTS TO A POISSON DISTRIBUTION

Probability limits	Observed value	Expected value	Chi-square
1.0–0.9	28	24	0.667
0.9–0.8	20	24	0.667
0.8–0.7	22	24	0.167
0.7–0.5	38	48	2.083
0.5–0.3	50	48	0.083
0.3–0.2	23	24	0.041
0.2–0.1	30	24	1.500
0.1–0.0	29	24	1.041
Total	240	240	6.250
Probability of Chi-square			0.5

The variability of such a series can be stabilized and the distribution made somewhat more normal by transformation to the square root of the count (actual number of eggs counted), Bartlett [2]. The mean square roots of the counts are listed in Table 1. The variances of the mean square roots were not significantly correlated to their means.

Analysis of variance of the transformed variates was made according to Snedecor [6], Section 11.8. The method of making two counts on each daily sample afforded an estimate of sampling error. The sheep-day interaction, Table 3, was the appropriate term for experimental error. In this experiment, the sampling error was not appropriate for use as the term for experimental error. Interaction was highly significant, denoting that σ_{12}^2 was not 0, thereby increasing the variability between sheep beyond that due to sampling errors alone. Where greater time intervals with seasonal fluctuations were involved [5, 7] interaction was probably much less—judging from graphs of the data. The significant interaction indicated that the changes in the daily environment of the sheep or worms, or both, did not influence the concentration of eggs in the feces of all lambs in a similar manner.

The pairing method would, therefore, not have been appropriate for these data.

TABLE 3

ANALYSIS OF VARIANCE OF EGG COUNTS, TRANSFORMED TO THE SQUARE ROOT

Source of variance	Sum of squares	Degrees of freedom	Mean square	Estimate of variances	
				Components	Estimate
Sheep	189.9033	11	17.2639*	$\sigma^2 + 2\sigma_{12}^2 + 40\sigma_1^2$	$\sigma_1^2 = .4093$
Days	76.9052	19	4.0476*	$\sigma^2 + 2\sigma_{12}^2 + 24\sigma_2^2$	$\sigma_2^2 = .1316$
Sheep × Days	186.0343	209	0.8901†	$\sigma^2 + 2\sigma_{12}^2$	$\sigma_{12}^2 = .2796$
Within Subcl.	79.4504	240	0.3310	σ^2	$\sigma^2 = .3310$

* Exceeds the .01 level using Sheep × Days as Exptl. Error.
† Exceeds the .01 level compared to Within Subclasses.

The variances for the 12 lambs were found to be homogeneous by Bartlett's test [6], Section 10.13. A standard difference, calculated from the error term, was thus appropriate for testing differences between means. The standard difference between two means, at the 5 percent level of probability was 0.585[2] eggs. Upon this basis, the mean counts of eight of the lambs were significantly higher than the lowest one.

While the Chi-square test of agreement to Poisson distribution on the basic data showed agreement, Coehran's test of the variance of the transformed variate showed non-agreement. Cochran [3] showed that $4x$ (within Subclasses sum of squares), using the square root transformation, should be distributed as Chi-square with Within Subclasses degrees of freedom. For these data the test of significance, Fisher [4], Section 16, is $\sqrt{2X^2} - \sqrt{2n-1}$ or $\sqrt{2\times4\times79.4} - \sqrt{2\times240-1}$ giving 3.3 which is highly significant. Even under conditions of this sort, Bartlett [2] has shown that the square root transformation is preferable if the variation of $\sqrt{x}$ is stable.

Several known sources of sampling error remain in the technique, but cannot be eliminated. The ability of any worker accurately to discover all the eggs on every slide will not be perfect. In addition, the volume of sample placed on the slide will not be exactly 0.1 ml. for every slide. These were probably the main sources of error which raised the variance out of agreement with Poisson.

Another source of increase in variance involves a correction for consistency of the feces, the factors being 1 for normally formed pellets, 1.5 for semi-mushy, 2 for mushy, and 4 for diarrhoeic feces. The

author's data were originally in this adjusted form, Table 1, giving an estimate of σ^2 of .518, and a Chi-square value of 9.7 with Cochran's test. Since the adjustment for moisture content of the feces made no major changes in relative standings of the lambs, it would appear that, for these data at least, this correction should not be made.

For many data on natural infestations, the additional correction for continuity, $\sqrt{x+0.5}$, instead of $\sqrt{x}$, would be necessary since the means are lower and counts of zero are not uncommon. While these transformations are time consuming, Bartlett [2] reminds the worker that "an unweighted average of the *observed* treatment response is not necessarily the best estimate of the *true* treatment response, and the average on the transformed scale will often be the better estimate when reconverted to the original scale."

REFERENCES

[1.] Ackert, J. E., L. L. Eisenbrandt, J. H. Wilmoth, B. Glading, and I. Pratt. "Comparative Resistance of Five Breeds of Chickens to the Nematode *Ascaridia lineata* (Schneider)." *Journal of Agricultural Research*, 50: 607–624 (1935).

[2.] Bartlett, M. S. "The Use of Transformations." *Biometrics*, 3: 1 (1947).

[3.] Cochran, W. G. "Some Difficulties in the Statistical Analysis of Replicated Experiments." *Empire Journal for Expe..mental Agriculture*, 6: 157 (1938).

[4.] Fisher, R. A. *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh (1940).

[5.] Gregory, P. W., R. F. Miller and M. A. Stewart. "An Analysis of Environmental and Genetic Factors Influencing Stomach Worm Infestation in Sheep." *Journal of Genetics*, 39: 391–400 (1940).

[6.] Snedecor, G. W. *Statistical Methods*, 4th Edition. Collegiate Press, Inc., Ames, Iowa (1946).

[7.] Stewart, M. A., R. F. Miller and J. R. Douglas. "Resistance of Sheep of Different Breeds to Infestation by *Ostertagia circumcincta*." *Journal of Agricultural Research*, 55: 923–930 (1937).

QUERIES

(46)

QUERY: In the formulas for missing values in experimental design, the error term is minimized. This of course inflates the F values and in turn the apparent significance of the effects and interactions under study. R. L. Anderson in the *Biometrics Bulletin*, Vol. 2, No. 3, gives formulas for removing this bias from the treatment sum of squares. Would it not be safer in all cases to minimize the sum of squares due to effects regardless of what this did to the error mean square?

This procedure has been used in one industrial experiment involving four factors by simply placing "Y" equal to the lost value and equating to zero the first derivative with respect to Y of the total attributed sum of squares, four main ones, six first order interactions and four second order interactions. This appeared to give a satisfactory replacement for the lost value in this isolated case.

What is the basis for preference in minimizing the residual mean square?

ANSWER: Yates has shown in the articles referred to in my discussion of missing plot techniques that if a missing value is estimated by minimizing the sum of squares for error (SSE) the resulting error mean square is an unbiased estimate of σ^2. Even though the treatment mean square is slightly inflated, the bias is known and can be removed if this is deemed advisable. It will be shown below that minimizing the sum of squares for treatments (SST) will make SST too small (and in many cases far too small) and SSE too large. I want to present a short example of how misleading it would be to minimize SST for two treatments. Suppose we had these plot yields for two treatments:

Treatment	Replication						Total
	1	2	3	4	5	6	
1	15	16	14	17	12	20	94
2	10	9	9	13	8	y	$49 + y$
Difference	5	7	5	4	4	$20 - y$	$45 + y$

The minimum SST is obviously 0 and this is obtained by setting $y = 45$. Hence we would have the first treatment 50% better than Treatment 2 in all five complete replications with the situation completely reversed by the missing plot in replication 6.

The statement that by minimizing *SST* we would be safer ignores the facts that:

(1) A test of significance should be as discriminating as possible, and a test based on a minimized *SST* (with resultant inflated *SSE* and far too small F) could be quite unreliable, as it is in the example cited above.

(2) Often the main use of the analysis of variance is to furnish a best unbiased estimate of σ^2. The method of estimating the missing value by minimizing *SSE* does not give an unbiased estimate of σ^2.

Another point would plague one if he estimated y by minimizing *SST*. How could he handle the missing degree of freedom? Since *SSE* has been inflated, it does not seem reasonable to reduce the degrees of freedom for error, and it would be hard to explain that we had one less degree of freedom for treatments. This problem could become quite perplexing for several missing values.

The foregoing discussion is clarified by considering the theoretical basis of the analysis of variance. The analysis of variance can be considered as a special application of normal regression theory. I will illustrate the general principles for a simple randomized block experiment with t treatments and r complete replications. The yield of any plot, y_{ij}, can be represented as

$$(1) \qquad y_{ij} = \mu + \tau_i + \rho_j + \epsilon_{ij} \qquad i = 1, \ldots, t; j = 1, \ldots, r,$$

where μ is some *fixed* over-all effect, τ_i is a *fixed* effect of the i-th treatment, ρ_j is a *fixed* effect of the j-th replicate, and ϵ_{ij} is the residual effect, unexplained by the three fixed effects. It is assumed that ϵ_{ij} are normally and independently distributed with mean 0 and common variance σ^2.

The best unbiased estimates of the fixed effects can be obtained by use of the method of least squares, whereby the residual sum of squares,

$$(2) \qquad \Sigma\Sigma\epsilon_{ij}^2 = \Sigma\Sigma(y_{ij} - \mu - \tau_i - \rho_j)^2,$$

is minimized with respect to each of these fixed effects, subject to the conditions,

$$(3) \qquad \Sigma\tau_i = \Sigma\rho_j = 0.$$

If μ, τ_i, ρ_j are replaced by their best estimates, m, t_i and r_j, respectively, in equation (2), an unbiased estimate of the sum of squares for error is obtained:

$$(4) \qquad SSE = \Sigma\Sigma(y_{ij} - m - t_i - r_j)^2.$$

The mean square for error, s^2, is an unbiased estimate of σ^2.

In order to test the null hypothesis that all the τ_i are 0, the original estimating equation (1) is replaced by

$$(1)' \qquad y_{ij} = \mu' + \rho'_j + \epsilon'_{ij}$$

and a new residual sum of squares SSE' is determined by minimizing $\Sigma\Sigma\epsilon'^2_{ij}$ with respect to μ' and the ρ'_j. The difference between SSE' and SSE is an unbiased estimate of the sum of squares for treatments, SST, given by the analysis of variance. SSE and SST are independent estimates of $\sigma^2\chi^2$; hence, the null hypothesis can be tested by means of the F ratio, where $F = \dfrac{SST}{t-1} / s^2$.

In the usual analysis of variance, SST is computed as

$$(5) \qquad \frac{\Sigma T_i^2}{r} - \frac{G^2}{rt},$$

where T_i is the total yield over all replications for the i-th treatment and G is the total yield for the entire experiment. This SST is unbiased and independent of SSE only when all treatments are included in each replicate. If any treatment has a missing observation, an unbiased SST, independent of SSE, can be obtained only by computing $SSE' - SSE$, which have been determined for only the existing data. Any substitute method must be evaluated as to how closely it can produce unbiased and independent estimates of SST and SSE.

The missing-plot technique of estimating the value of a missing plot, say y_{11}, by minimizing SSE furnishes an SST biased slightly upwards (by $\dfrac{t}{t-1} \cdot t_1^2$). However, we see that SSE is unbiased. The residual sum of squares is

$$(6) \qquad \Sigma\Sigma'(y_{ij} - m - t_i - r_j)^2 + (y_{11} - m - t_1 - r_1)^2,$$

where $\Sigma\Sigma'$ extends over only the existing data. When this residual is minimized with respect to y_{11}, it is seen that $y_{11} = m + t_1 + r_1$; hence, the residual sum of squares is unchanged and is exactly SSE.

If y_{11} were estimated by minimizing SST from equation (5), it can be shown that

$$y_{11} = m + t_1 + r_1 - \frac{rt}{t-1} \cdot t_1.$$

Applying this result to (6), we find that in this case SSE is inflated by $\left[\dfrac{rt}{t-1} \cdot t_1\right]^2$. Similarly, SST is too small by $\dfrac{(r-1)t}{t-1} \cdot t_1^2$.

For those who wish to use the method of covariance, I might mention that in this case estimating equations (1) and (1)′ merely have a new variable y_{11} added. The regression method then would adjust all sums of squares for y_{11}, so that unbiased and independent values of SST and SSE would be determined.

It should be pointed out that these results could equally well be obtained by use of the method of maximum likelihood, and the general conclusions given here can be extended to more complicated experimental designs.

R. L. ANDERSON

(47)

QUERY: The following problem has worried me for some time. In almost all modern textbooks of statistics the procedure given for comparing two samples as to mean and variance involves the following:

Consider two samples of N_1 and N_2 variates respectively; we wish to know whether they can have come from the same population. The squares of the deviations of the variates of each sample from its own mean are added together and divided by N_1+N_2-2. This is considered an estimate of the variance (s^2) of the population from which both samples were drawn, on the null hypothesis that they were in fact from a common population. Why is it not more accurate to obtain such an estimate of s^2 as follows: take the grand mean of all variates in both samples, take deviations from it, square, add, and divide by N_1+N_2-1? That would seem to be the obvious way to get a best estimate of combined variance, and usually it will lead to larger estimates than will the conventional method.

ANSWER: Apparently you do not distinguish clearly two different problems, estimation and testing hypotheses. Perhaps also you have not discriminated among the circumstances in which each problem is to be solved.

Consider first the estimation of variance. You have given the correct method if the two samples are assumed to have been randomly drawn from a common normally distributed population; but I see no reason to think that in this case your estimate "usually will lead to larger estimates than will the conventional method." This conventional method is appropriate if the sampled populations are assumed to have the same σ^2 but not necessarily the same means. The corresponding analysis of variance is this:

Source of Variation	Degrees of Freedom	Mean Square
Means	1	M_M
Individuals	N_1+N_2-2	M_I
Total	N_1+N_2-1	M_T

M_I is here an estimate of σ^2; it is a weighted mean of the estimates given by the two samples, the assumption being that they are estimating a common population variance. M_M is now an estimate of $\sigma^2 + k_0\sigma_M^2$, where k_0 is $N_1+N_2-\frac{N_1^2+N_2^2}{N_1+N_2}$, and where σ_M^2 is the component of variance introduced by the difference between the population means. Under these conditions your estimate, M_T, does not estimate σ^2 but will usually be larger than σ^2 because of the component σ_M^2. This is the situation you probably had in mind in making the last statement in your query.

Let us now look at the conventional test of significance, $F = M_M/M_I$; that is, the test of the null hypothesis, $\sigma_M^2 = 0$, under the assumption that σ^2 is the variance common to the sampled populations. An essential condition is that, under the null hypothesis, M_M and M_T shall be independent and shall both be estimates of σ^2. If you used M_T (your proposed estimate of variance) in the denominator, the resulting ratio would not be distributed as F; moreover, it seems obvious that the test corresponding to your proposal would be less sensitive than F.

For the t-test a similar argument can be constructed, the principles being the same.

J. G. DARROCH

(48)

QUERY: I have k samples of r with varying sample size. I wish to test the hypothesis that they are random samples from a common normal population. Using Fisher's z-transformation, the quantity,

$$(1) \qquad \Sigma N z^2 - \frac{(\Sigma N z)^2}{\Sigma N},$$

is said to be distributed as chi-square with $(k-1)$ degrees of freedom. I have tried to derive this by taking

$$(2) \qquad \chi^2 = \Sigma(o-e)^2/e,$$

but without success. Nor have I been able to derive the formula from

$$(3) \qquad \chi^2 = (\Sigma x^2)/\sigma^2.$$

It seems to me that in (3) we have only the numerator of formula (1). That is, unless we can assume that z is normally distributed with unit standard deviation. Would you please put me on the right track?

ANSWER: Formulas (1) and (3) are seen to be equivalent if the following pertinent facts are noted. (i) The quantity z is almost normally distributed with variance $1/(n-3)$ where n is the size of the sample (R. A. Fisher, *Statistical Methods,* Section 35). Presumably your $N = n-3$. (ii) The variate x is normally distributed with variance σ^2; hence $\sigma^2 = 1/(n-3)$. (iii) Your formula (1) is identical with

$$\Sigma N(z-\bar{z})^2 = \Sigma(n-3)(z-\bar{z})^2 = \Sigma(z-\bar{z})^2/\sigma^2,$$

the weighted sum of squares of deviations of z. This quantity is distributed as χ^2, with $(k-1)$ degrees of freedom, under the hypothesis that the z's are all estimates, with the above variances, of a single parameter (R. A. Fisher, *Statistical Methods,* Section 21.03).

Formula (2) is an approximation to χ^2, applicable to enumeration data. For a discussion, see M. G. Kendall, *The Advanced Theory of Statistics,* Vol. 1, Chapter 12.

OSCAR KEMPTHORNE

(49)

QUERY: I am interested in the problem of heredity as an etiological factor in homosexuality.

It was shown by Goldschmidt in his genetic study of butterflies that he could establish successive stages of intersexuality from the full male through slightly and markedly feminized males to the completely transformed female sex intergrade. He also produced a set of intersexes ranging from complete female through all intermediate stages to the male sex intergrade. It cannot be assumed that these results apply directly to men but to explain intersexual stages and the transformation of sex many geneticists have felt that sex is determined by sex chromosomes and also by the potentialities of the autosomes.

It has been shown that changes in temperature, nutrition, light, hormones, vitamins, and changes in oxidation rates of anabolism and catabolism in organisms (changes in physical and chemical environment) may modify the sex of a developing organism to change the somal configuration of the organism.

If the developmental balance theory of Goldschmidt is correct then intersexes are all genetically of the female type. If this is true then all of the descendants of intersexes would be female. There also should be more males found among the siblings of intersexes. If it is

true that a percentage of male homosexuals are real male sex intergrades which are genetically female, but which have lost all morphological sex characteristics except their chromosome formula, then an undue preponderance of males should be found among the siblings of male homosexuals. The normal ratio of 106 males: 100 females at birth (including live births and stillbirths, excluding stillbirths of unknown sex) should then be altered to show an increase in males since a series of genotypical females would be among the phenotypical male probands.

I am enclosing the data as determined for two groups. I have attempted to determine the significance of the difference in the sex ratio, but am not satisfied that this method is appropriate to the problem or that I am handling it correctly. Does this method determine whether or not the differences are significant? What is the standard error of the sex ratio?

Group 1. $N = 91$, males = 48, theoretical sex ratio = 106, how reasonable to get 111?

$$p = 48/91 = 0.52,\ q = 48$$

$$\sigma = \sqrt{\frac{(0.52)\,(0.48)}{91}} = 0.05$$

$$\frac{0.52 - 0.50}{0.05} = 0.4 \text{ times } \sigma, \text{ which is not 3 times}$$

the σ, therefore not significant.

Group 2. $N = 148$, males = 82, theoretical sex ratio = 106, how reasonable to get 124?

$$p = 106/148 = 0.71,\ q = 1 - 0.71 = 0.29$$

$$\sigma = \sqrt{(148)\,(0.71)\,(0.29)} = 5.52.$$

The difference between 106 and 124 is more than 3 times σ, and therefore arises other than by chance.

ANSWER: In your first group there are two corrections that should be made. First, in the formula for σ, $p = 106/206 = 0.515$, the *theoretical* probability, instead of the sample estimate which you have apparently indicated. Second, in the difference between the theoretical and sample probabilities you have put the theoretical value equal to 0.5 instead of the correct value shown above. The calculations for group 1 are as follows:

$$p = 0.515,\ q = 0.485,\ \sigma = \sqrt{(0.515)\,(0.485)/91} = 0.052$$

Sample ratio = 48/91 = 0.527, $t = (0.527 - 0.515)/0.052 = 0.23$

In your group 2 the computations seem to have no relation to the facts and have led you to a false conclusion. Here,

$p = 106/206 = 0.515$, as before, and $\sigma = \sqrt{(0.515)(0.485)/148} = 0.041$

Sample ratio = 82/148 = 0.554, $t = (0.554 - 0.515)/0.041 = 0.95$

As in group 1, the deviation of the sample ratio is not so large as σ.

The sex ratio as you define it has the standard error,

$$\sigma = \frac{1}{1+\rho}\sqrt{\frac{\rho}{n}},$$

where ρ is the hypothetical sex ratio (in your case, $\rho = 1.06$) and n is sample size.

In groups like yours it may be slightly easier to compute t directly:

$$t = \frac{f\rho - m}{\sqrt{n\rho}},$$

where f is the number of females and m the number of males in the sample.

GEORGE W. SNEDECOR

(50)

QUERY: A problem has arisen in developing a deer browse sampling technique to appraise the available browse supply in deer wintering areas in the Adirondack region of New York State and to establish a reasonably accurate method of ascertaining the yearly change in the browse supply. Because of the extent of the wintering areas, they range from 1½ to 14 square miles in area, it is not feasible to use random samples based on a grid system since the location and appraisal of each plot would entail the use of surveying instruments to accurately spot a number of scattered plots.

The method we now employ is as follows: two or three traverse lines are drawn at right angles to the main stream bed or valley of each area. The lines extend to the perimeters of each area. This insures that all altitudinal regions within each area will be represented in approximately the same proportion as they exist throughout the entire area. Each traverse line is then divided into 100 equal parts and numbered 1 to 100. Fifty random numbers are selected from Tippett and the numbers drawn are used to spot plots on the traverse line, the plots being irregularly spaced due to random selection. The points selected are staked and the area within a 25-foot radius comprises the plot to be studied. All browse species within each plot are listed on a field

schedule and the relative density of each species is noted. Since the method employed is a hodge-podge of selective and random sampling (selective in choosing traverse lines and random in locating the plots along the line) the question arose as to whether it would not be better to use selective sampling for both traverse lines and plots along the lines. The selection of the traverse line violates the principle of random sampling and thus renders the data gained unfit for interpretative analysis inasmuch as the standard deviation would be meaningless. On the other hand randomization of the plots along the line would enable one to calculate the standard deviation and standard error but since these statistics would be vitiated by the selection originally used in locating the traverse line it is questionable as to whether any useful purpose is served in using a random sample of plots along the line.

I will appreciate your views on this subject and will welcome any suggestion you may make or any books you may know that treat of similar problems. I wish to point out in closing that the main purpose of the study is to determine yearly changes in the browse supply. An accurate appraisal of the browse supply, or a measurable degree of accuracy, on each area is desirable, but probably unattainable without random sampling.

ANSWER: Your two objectives (a) to appraise the available browse supply and (b) to ascertain the yearly change in browse supply, indicate the general field procedure known as the method of double sampling.

For the moment consider objective (a) above. Your present method of running traverse lines at right angles to the main stream, from ridge to ridge—that is, across the stratification which is roughly parallel to the stream—seems efficient and practical from this distance. Right here I should like to suggest that the number of sample points along the traverses (centers of sample-plot circles of 25-ft. radii) be proportional to length of traverse—such, for instance, as one in each 100 feet of traverse length, taken systematically.

If you are required to appraise the browse supply of *each* wintering area with satisfactory precision, I imagine that 2 or 3 traverse lines as now employed are insufficient. Perhaps a random sample of 2 traverses to each mile of distance, running at right angles to an imaginary line from the down-stream end to the top of the valley, would prove workable. This scheme implies the division of a wintering area into

blocks having one dimension of one mile, but the other dimension varying with distance from base line to ridge. It is discussed in detail in Chapter VII of "Sampling Methods in Forestry and Range Management" by Schumacher and Chapman (Bulletin 7, School of Forestry, Duke University, 1942).

On the other hand, if you are required to appraise the supply of only the aggregate of all areas, and not of each individual area, the field work may be restricted considerably, perhaps to your present scheme of 2 or 3 traverses in each wintering area. On the supposition that you require the sampling error of your appraisal, a random sample of traverses should be taken from each area.

The above discussion has to do with objective (a) only—that is, the appraisal of the browse supply of the particular year. The adjustment to the supply of the succeeding year should require relatively little field work. For this purpose, the circular plot of 25-ft. radius should serve as a satisfactory sampling unit. If in the first year a certain (relatively small) number of these are identified by a permanent post at their centers, and in the second (and each succeeding) year, the browse supply on these permanent plots alone be measured, then the regression of the second-year browse on the first-year browse of these plots is the adjusting equation required. One inserts into the equation the average browse supply (on the circular plot basis) as obtained in the comparatively heavy sampling of the first year and solves for the corresponding value of the second year.

It thus seems that a somewhat heavy sampling need be made only once every several years—that is, only at such time as the regression of succeeding years' measurements of individual plots on those of the base year no longer serve as satisfactory adjustment.

F. X. SCHUMACHER

ABSTRACTS*

(40)

GREENWOOD, MARY L. (University of Connecticut). **Studies on the Palatability of Kale in Relation to Variety and Cooking Procedure.**

The relative palatability was tested by means of consumer preferences. The ranks were converted to mean normal deviates. These transformed scores were then compared in a series of analyses of variance. Where a larger number of alternatives had to be compared than could be handled conveniently at one time, the experiments were arranged in balanced incomplete blocks. The method of ranks seems adapted to relatively complex experiments for measuring food preferences.

(41)

BROWN, GEORGE W. (Iowa State College). **Discriminant Functions.**

The logical foundations of discriminant function techniques are developed from an elementary point of view. The basic model postulates underlying "measurements" of the form

$$\xi_1 = \alpha_1 + \beta w + \varepsilon_1$$
$$\xi = \alpha_2 + \varepsilon_2$$
$$\vdots$$
$$\xi_p = \alpha_p + \varepsilon_p$$

where the α_i and β are unknown constants, ε_i are independent errors of measurement, and w is a parameter varying over a set of populations. In addition, it is postulated that the observables are of the form

$$x_m = \sum_{n=1}^{p} l_{mn}\,\xi_n,\ m = 1, \ldots, p$$

where the $\{l_{mn}\}$ are unknown coefficients. If $\{\sigma_{mn}\}$ is the matrix of variances and covariances of the x's, and $\{\sigma^{mn}\}$ is the inverse of this matrix, and if b_m is the regression coefficient of x_m on w, then the

* All papers read at the joint meetings of the Biometrics Section and the Institute of Mathematical Statistics held in conjunction with the 113th Annual Meeting of the American Association for the Advancement of Science in Boston, December 27–29, 1946.

function $X = \sum_{mn=1}^{p} \sigma^{mn} b_m x_n$ is proportional to the hypothetical measurement ξ_1, the only one of the ξ_n which is related to w. This X discriminates among w values to the same extent that ξ_1, were it available, would discriminate. X is, of course, also that linear combination (unique except for a constant of proportionality) of x's for which the ratio of the mean square for regression to the mean square for error is a maximum, or what is the same thing, X is the linear combination of x's which has greatest correlation with w. The coefficients of X are proportional to those which would have been obtained from a formal regression analysis of w on $x_1, x_2, \ldots, x_p$. The usual procedure for testing against subsets of the x's still holds in this situation, in spite of the inversion of roles of dependent and independent variables.

Two generalizations to multi-dimensional population parameters are possible. Let $w_1, w_2, \ldots, w_s$, be parameters associated with a set of populations. The case $s = 2$ is considered. Let there be hypothetical measurements $\xi_1, \xi_2, \ldots \xi_p$, with

$$\begin{aligned}
\xi_1 &= \alpha_1 + \beta_1 u + \gamma_1 v + \varepsilon_1 \\
\xi_2 &= \alpha_2 + \beta_2 u + \gamma_2 v + \varepsilon_2 \\
\xi_3 &= \alpha_3 + \varepsilon_3 \\
&\vdots \\
\xi_p &= \alpha_p + \varepsilon_p
\end{aligned}$$

Let b_m be the regression coefficient of x_m on u, c_m the regression coefficient of x_m on v, and let σ_{mn} be the variance-covariance matrix among the $\{\varepsilon_m\}$. Then $X_1 = \sum \sigma^{mn} b_m x_n$ and $X_2 = \sum \sigma^{mn} c_m x_m$ are independent linear combinations of ξ_1 and ξ_2 alone, so that x_1 and X_2 together contain all the information in the x's relevant to u and v. X_1 and X_2 together yield the best 2-dimensional description of the populations. The other generalization corresponds to finding the best 1-dimensional description of the populations, leading to the solution of a determinantal equation. X_1 and X_2 are the 2-dimensional discriminant function; the best 1-dimensional linear function, in the sense of ratio of mean square for regression to mean square for error, is the "reduced" discriminant function. In general, when $s > 2$, there is a hierarchy of reduced discriminant functions with dimensionality $1, 2, \ldots, s-1$.

NEWS AND NOTES

Statisticians and those interested in statistics from the United States and foreign countries will convene at Virginia Polytechnic Institute, Blacksburg, Virginia, for the Statistical Summer Session, August 5 to September 5. This Summer Statistical Session is being sponsored jointly by Virginia Polytechnic Institute, University of North Carolina, University of Michigan, Iowa State College, and the Federal Bureau of Agricultural Economics. The faculty will consist of: WALTER A. HENDRICKS, B.A.E., U.S.D.A.; RENSIS LIKERT, University of Michigan; H. L. LUCAS, University of North Carolina; MAURICE G. KENDALL, England; GEORGE W. SNEDECOR, Iowa State College; FRANK YATES, Rothamsted Experiment Station, England; EARL E. HOUSEMAN, B.A.E., U.S.D.A.; RAYMOND J. JESSEN, Iowa State College; and BOYD HARSHBARGER, Virginia Polytechnic Institute. The following courses will be offered for credit: Engineering Statistics; Statistical Methods; Design of Animal Experiments; Schedule Design and Interview Techniques for Sample Surveys; Sampling Design and Analysis; Mathematical Theory of Sampling; Seminar; Mathematical Statistics; and Experimental Design. In addition to the faculty, probable Seminar speakers are: W. F. CALLENDAR, W. G. COCHRAN, GERTRUDE M. COX, W. E. DEMING, GEORGE GALLUP, MORRIS HANSEN, HAROLD HOTELLING, ARNOLD KING, and CHARLES F. SARLE. Inquiries regarding the Summer Session should be addressed to Boyd Harshbarger, Professor of Statistics, Statistical Summer Session, Virginia Polytechnic Institute, Blacksburg, Virginia. . . . The Biological Methods Group of the Society of Public Analysts and Other Analytical Chemists held its annual meeting on December 16, 1946. Two papers were read: 1. "The assay of anti-thyroid substances using tadpoles" by H. M. BRUCE. In this paper she was concerned almost entirely with matters of biological technique. 2. "The computation of microbiological assays of amino-acids and other growth factors" by ERIC B. WOOD. He points out that microbiological assays which did not show linear relationship between dose and response are nearly always linearized when the logarithms of both the dose and the response are taken as metameters. . . . R. HOWARD PORTER, Director, Instituto Agronómico Nacional, left Iowa State College a year ago for Paraguay to work on the Food Production Program. His work is primarily concerned with varietal testing of many kinds of crops and seed production of the

improved varieties which have been tested in the past. In addition, he is doing work with seed disinfectants and agronomic experiments such as spacing and fertility tests. There is a nursery for testing many kinds of vegetables for their adaptation to this region. . . . HOWARD L. STIER has returned to the University of Maryland as Professor of Marketing and head of the department after more than five years in the army. He had been a member of the Department of Horticulture at the University before entering the army in 1941. During the war, Dr. Stier served as Chief Statistician of the Quartermaster Corps and later as Director of the Statistics and Progress Reporting Division of War Assets Administration, rising from the rank of 1st Lieutenant to Colonel. He was awarded the Legion of Merit for "outstanding service" in the Quartermaster Corps. . . . BURCH H. SCHNEIDER, Professor of Animal Husbandry, West Virginia University, Morgantown, deserves a word of commendation for his unlimited patience in the compilation of the material for "Feeds of the World." . . . T. A. BANCROFT, now at the University of Georgia, Athens, will go to Alabama Polytechnic Institute at Auburn the middle of June. He goes to organize and direct a new statistical laboratory under the sponsorship of the Dean of the Graduate School and the Alabama Polytechnic Institute Research Council. The functions of the Statistical Laboratory are: (1) to provide instruction in statistical methods on the senior and graduate level, (2) to provide a consultation service for research scientists at the university and in the state, and (3) to provide computational service for college and state agencies. Both the State Agricultural and Engineering Experiment Stations are located at Auburn. ALONZO COHEN, now working in mathematical statistics at the University of Michigan, will take his place at the University of Georgia. . . . MARGARET E. RICE has resigned her position as statistician with the TVA at Wilson Dam to become supervisor of public health statistics at Mississippi State Board of Health at Jackson. She assumed her new position February 1. . . . The Tennessee Public Health Association held its annual meeting in Nashville, May 5–7. ANN DILLON arranged the program. . . . W. EDWARDS DEMING, the man who travels to England, India, and Japan, should provide some highlight reports for us, especially when he writes, "I had a wonderful time in India and came away with many puzzling thoughts." . . . E. J. WILLIAMS, Division of Forest Products, Council for Scientific and Industrial Research, South Melbourne, Australia, writes: "Since I began work here at the

Division of Forest Products some six years ago, the statistical work required by the Division has increased considerably. There are now two of us on the statistical staff, with five computers. Much of the work done here is routine, such as computing of test results, and analyses of variance. However, we get a number of interesting problems. Timber being subject to defect and loss, the analysis of experiments involves quite a lot of treatments of non-orthogonal data, for which we have been developing some useful approximate methods, as well as iterative methods for arriving at the exact solution. We have come across a number of problems in design, such as factorial experiments, which have to be fitted into unbalanced incomplete block designs, so as to confound only interactions. On the theoretical side, I have been playing round with significance tests in multivariate analysis, but so far do not see any exact approach. R. A. FISHER, I believe, has been delving into the problem. You will be interested to hear we are making use of 7×7 Latin square designs in studying the effect of two consecutive processes in laboratory sheet making; time of drawing of stock, and position in the press. By randomizing the sheets in such a way that each sheet from one time of drawing occurred at a different one of the seven positions in the press, in successive replications, we were able to disentangle the two effects satisfactorily." He inquired whether we have met David Duncan. We have. The Editorial Committee would be glad to have others send us brief reports regarding their work. . . . H. J. MILLER, Assistant Professor of Plant Pathology at Pennsylvania State College, comments on this bulletin: "For the average biological worker, it affords an opportunity to have questions answered by competent statisticians such as are not available in many of our institutions." . . . C. H. GOULDEN, Officer in Charge, Cereal Division, Department of Agriculture, Winnipeg, Canada, writes: "I should like to take this opportunity of stating that the *Biometrics Bulletin* is filling a very definite need. It is not possible for most of us to read more than one half of the papers that are being published at the present time on statistics. There must be some journal that either digests articles from other journals or presents general articles by well-known writers on particular phases of statistical methods." Mr. Goulden thinks our new format is a real improvement. We hope you agree.